Eur päisches Patentamt

European Patent Office

Office eur péen des brevets

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. **Patent application No.** **Demande de brevet n°**

02368090.3

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

DEN HAAG, DEN
THE HAGUE, 03/10/02
LA HAYE, LE

Europäisches Patentamt
European Patent Office
Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 02368090.3
Demande n°:

Anmeldetag:
Date of filing: 22/08/02
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
INTERNATIONAL BUSINESS MACHINES CORPORATION
Armonk, NY 10504
UNITED STATES OF AMERICA

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Method and system for splitting and sharing routing information between several routers acting as a single border router

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:

/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

METHOD AND SYSTEM FOR SPLITTING AND SHARING ROUTING INFORMATION BETWEEN SEVERAL ROUTERS ACTING AS A SINGLE BORDER ROUTER

Technical field of the invention

The present invention is directed to computer networks, and more particularly to a method and system, in an Internet Protocol (IP) network, for splitting and sharing Border Gateway Protocol (BGP) routing information between several routers acting as a single border router.

Background art

Internet

The Internet is a global network of computers and computers networks (the "Net"). The Internet connects computers that use a variety of different operating systems or languages, including UNIX, DOS, Windows, Macintosh, and others. To facilitate and allow the communication among these various systems and languages, the Internet uses a language referred to as TCP/IP ("Transmission Control Protocol/Internet Protocol"). TCP/IP protocol supports three basic applications on the Internet :

- transmitting and receiving electronic mail,
- logging into remote computers (the "Telnet"), and
- transferring files and programs from one computer to another ("FTP" or "File Transfer Protocol").

TCP/IP

The TCP/IP protocol suite is named for two of the most important protocols:

- a Transmission Control Protocol (TCP), and
- an Internet Protocol (IP).

Another name for it is the Internet Protocol Suite. The more common term TCP/IP is used to refer to the entire protocol suite. The first design goal of TCP/IP is to build an interconnection of networks that provide universal communication services: an

"internetwork", or "internet". Each physical network has its own technology dependent communication interface, in the form of a programming interface that provides basic communication functions running between the physical network and the user applications. The architecture of the physical networks is hidden from the user. The second goal of TCP/IP is to interconnect different physical networks to form what appears to the user to be one large network.

TCP is a transport layer protocol providing end to end data transfer. It is responsible for providing a reliable exchange of information between 2 computer systems. Multiple applications can be supported simultaneously over one TCP connection between two computer systems.

IP is an internetwork layer protocol hiding the physical network architecture bellow it. Part of the communicating messages between computers is a routing function that ensures that messages will be correctly directed within the network to be delivered to their destination. IP provides this routing function. An IP message is called an IP Datagram.

Application Level protocols are used on top of TCP/IP to transfer user and application data from one origin computer system to one destination computer system. Such Application Level protocols are for instance File Transfer Protocol (FTP), Telnet, Gopher, Hyper Text Transfer Protocol (HTTP).

World Wide Web

With the increasing size and complexity of the Internet, tools have been developed to help find information on the network, often called navigators or navigation systems. Navigation systems that have been developed include standards such as Archie, Gopher and WAIS. The World Wide Web ("WWW" or "the Web") is a recent superior navigation system. The Web is :

- an Internet-based navigation system,
- an information distribution and management system for the Internet, and
- a dynamic format for communicating on the Web.

The Web seamlessly, for the use, integrates format of information, including still images, text, audio and video. A user on the Web using a graphical user interface ("GUI", pronounced "gooey") may transparently communicate with different host computers on the system, and different system applications (including FTP and

Telnet), and different information formats for files and documents including, for example, text, sound and graphics.

IP Router

A "Router" is a computer that interconnects two networks and forwards messages from one network to the other. Routers are able to select the best transmission path between networks. The basic routing function is implemented in the IP layer of the TCP/IP protocol stack, so any host (or computer) or workstation running TCP/IP over more than one interface could, in theory, forward messages between networks. Because IP implements the basic routing functions, the term "IP Router" is often used. However, dedicated network hardware devices called "Routers" can provide more sophisticated routing functions than the minimum functions implemented in IP.

Routing Protocol

When data is sent to a remote destination, each IP datagram is first sent to a local router. An incoming datagram that specifies a destination IP address other than one of the local router IP address is treated as a normal outgoing datagram. This outgoing datagram is subject to the IP routing algorithm of the router, which selects the next hop for the datagram. The router forwards each datagram towards its final destination. A datagram travels from one router to another until it reaches a router connected to the destination. Each intermediate router along the end-to-end path selects the next hop used to reach the destination. The next hop represents the next router along the path to reach the destination. This next router can be located on any of the physical networks to which the intermediate router is attached. If it is a physical network other than the one on which the host originally received the datagram, then the result is that the intermediate router has forwarded the IP datagram from one physical network to another.

An "IP routing table" in each router is used to forward datagrams between networks. A basic IP routing table comprises information about the locally attached networks and the IP addresses of other routers located on these networks, plus the networks they attach to. A routing table can be extended with information on IP networks that are farther away, and can also comprise a default route, but it still remains a table with limited information. A routing table represents only a part of the whole IP

networks. A router having such a routing table is called "a routers with partial routing information".

A robust routing protocol must provide the ability to dynamically build and manage information in the IP routing table. As the changes in the network topology may occur, the routing tables must be updated with minimal or without manual intervention.

IP Addressing

IP addresses are used by the IP protocol to uniquely identify a host on the Internet. Strictly speaking, an IP address identifies an interface that is capable of sending and receiving IP datagrams. Each IP datagram (the basic data packets that are exchanged between hosts) comprises a source IP address and a destination IP address. IP addresses are represented by a 32-bit unsigned binary value which is usually expressed in a dotted decimal format. For example, 9.167.5.8 is a valid Internet address. An IP address is divided between a network and a host part, the first bits of the IP address specifying how the rest of the address is divided. The mapping between the IP address and an easier-to-read symbolic name, for example myhost.ibm.com, is done by the "Domain Name System" (DNS).

Subnetworks

An IP address comprises a list four numbers, each number in a range from 0 to 255 and separated by a point "." All possible addresses in the IP network (in other word the entire IP address space) are between address 0.0.0.0 and address 255.255.255.255. The total number of available IP addresses is 2 power 32. An IP address can also be represented by converting each digit in a binary value. For instance, the IP address "0.1.2.3" can be represented by the binary value : "00000000 00000001 00000010 00000011". The binary representation is especially important due to several of its properties :

- An address which is the result of an operation consisting in adding 1 to the binary representation of a first address is considered as "contiguous" to said first address.

- The total address space can be split into smaller blocks of contiguous addresses using a binary mask. These blocks of contiguous addresses are called "subnetworks" or "subnets". A subnet is defined by two values :
  - the size of the mask to apply (a number “n” between 0 and 32), and
  - the first address of the block (which must be a multiple of 2 power (32-n)).

Once these two values are given, it is very simple to check if an address ”x” belongs to this subnet : the binary mask is applied on the binary representation of the IP address “x”. The result must be equal to the first address of the subnet. The mask is applied to an address by means of a logical "AND" operation between the bits of the mask and the corresponding bits of this address. As the mask is a characteristic of the subnet, it is usually called “subnet mask”. The subnet mask is entirely defined by its length "n" currently denoted “/n”.

For instance, the subnet defined by the subnet mask /30 and the address 1.1.1.0 consists in a block of contiguous addresses between 1.1.1.0 and 1.1.1.3, with the four addresses 1.1.1.0, 1.1.1.1, 1.1.1.2 and 1.1.1.3. In other words mask /30 can be represented by the binary string : "11111111 11111111 11111111 11111100". An address belongs to the subnet previously defined if, when the binary mask /30 is applied on the binary representation of the address, the result is equal to the first address of the subnet, in the present case : "00000001 00000001 00000001 00000000". For instance, if the mask /30 is applied to the address "1.1.1.2" represented by the string "00000001 00000001 00000001 00000010", the result is equal to "00000001 00000001 00000001 00000000" which is the first address of the subnet. Therefore the address "1.1.1.2" belongs to the subnet. On the contrary, if the same operation is done on address "2.2.2.2" represented by the binary string "00000010 00000010 00000010 00000010", the result is "00000010 00000010 00000010 00000000". The result is different from the first address of the subnet. Therefore, address “2.2.2.2“ does not belong to the subnet.

Autonomous System (AS)

A more intelligent routers are required if :

- The router has to know routes to all possible IP networks.
- The router has to have dynamic routing tables, which are kept up-to-date with minimal or no manual intervention.

- The router has to be able to advertise local changes to other routers.

Advanced forms of routers use additional protocols to communicate with each other. A number of protocols of this kind exist. For instance, using the internet terminology, there is a concept of a group of networks, called an "Autonomous System" (AS) which is administered as a unit. Routing within an Autonomous System (AS) and routing outside an Autonomous System (AS) are treated as different issues and are addressed by different protocols.

An Autonomous System (AS) is defined as a logical portion of a larger IP network. an AS normally comprises an internetwork within an organisation. It is administered by a single management authority. An AS can connect other ASs managed by the same organisation. It can also connect other public or private networks.

Some routing protocols are used to determine the routing path within an AS. Others are used between a plurality of ASs :

- Exterior Gateway Protocols (EGPs) allow the exchange of summary information between separately administered ASs. An example of this type of routing protocol is the Border Gateway Protocol (BGP) also called EBGP for Exterior Border Gateway Protocol.
- Interior Gateway protocols (IGPs) allow routers to exchange information within an AS. Examples of these protocols are Open Short Path First (OSPF) and Routing Information Protocol (RIP). The routing protocol BGP can also be used within an AS as IGPs. In this particular case BGP is called Internal Border Gateway Protocol (IBGP).

Exterior Gateway Protocol (EGP)

The Exterior Gateway Protocol (EGP) has a particular historical merit. it was one of the first protocols developed to communicate between Autonomous Systems (ASs). This protocol is described in RFC 904. EGP assumes that the network comprises a single backbone and a single path exists between any two autonomous systems. Due to this limitation, the current use of EGP is minimal. In practice, EGP has been progressively replaced by BGP. EGP is based on periodic polling using hello/I-hear-you message exchanges, to monitor neighbour reachability and poll requests to solicit update responses. Exterior gateways connected to an AS can advertise only those destinations networks reachable entirely within that gateway's AS. An exterior gateway using EGP passes along information to its EGP neighbours

but does not advertise reachability information about its EBGP neighbours (gateways are neighbours if they exchange routing information) outside the AS. The routing information from inside an AS must be collected by this EGP gateway, usually via an Interior Gateway Protocol (IGP).

Border Gateway Protocol (BGP)

The Border Gateway Protocol (BGP) is an Exterior Gateway Protocol. It was originally developed to provide a loop-free method for exchanging routing information between Autonomous Systems (ASs). BGP has since evolved to support the aggregation and summarisation of routing information. BGP is an IETF draft standard protocol described in RFC 1771. The version described in this RFC is BGP version 4 (BGP-4).

A system running the Border Gateway Protocol (BGP) is called a "BGP speaker". A pair of BGP speakers exchanging inter AS routing information are called "BGP neighbours". BGP neighbours can be of two types :

- Internal : a pair of BGP neighbours in the same Autonomous System. Internal BGP neighbours must present a consistent image of the AS.
- External : a pair of BGP neighbours in different Autonomous Systems. External BGP neighbours must be connected by a BGP connection.

A "BGP session" is a TCP session established between BGP neighbours exchanging routing information using BGP. The neighbours monitor the state of the session by sending a "keepalive" message regularly. A "border router" or "border gateway" is a router that has a connection to multiple Autonomous Systems.

The IP address of a border router is specified as a next hop destination when BGP advertises an AS path (list of the As numbers traversed by a route when exchanging routing information) to one of its external neighbours. Next hop border routers share a physical connection with both the sending and receiving BGP speakers.

BGP defines two types of connections :

- Physical connection : an AS shares a physical network with another AS, and this network is connected to at least one border router from each AS. Since these two routers share a network, they can forward datagrams to each other without requiring any inter AS or intra AS routing protocols.
- BGP connection : a BGP connection means that there is a BGP session between a pair of BGP speakers, one in each AS. This session is used to communicate the routes through the physically connected border routers that can be used for specific networks. BGP requires that the BGP speakers must be on the same network as the physically connected border routers so that the BGP session is also independent of all inter AS or intra AS routing protocols. The BGP speakers do not need to be border routers and vice versa.

Note : the term BGP connection can be used to refer to a session between two BGP speakers in the same AS.

Routing policies are not defined in the BGP protocol but are selected by the AS authority and presented to BGP in the form of implementation specific configuration data. Each BGP speaker must :

- evaluate different paths to a destination from the border router(s) for an AS connection,
- select the best one that complies with the routing policies in force, and
- advertise that route to all of its BGP neighbours at that AS connection. Rather that exchanging simple metrics counts, BGP communicates entire paths to its neighbours.

BGP determines a preference order by applying a function mapping each path to a preference value and selects the path with the highest value.

BGP only advertises routes that it uses itself to its neighbours. That is, BGP conforms to the normal Internet hop-by-hop paradigm, even through it has additional information in the form of AS paths and theoretically could be capable of informing a neighbour of a route it would not use itself. When two BGP speakers form a BGP session, they begin by exchanging their entire routing tables. Routing information is exchanged via "update" messages. In addition to the reachability and next hop information, the routing information contains the complete AS path to each listed destination in the form of a list of AS numbers. After BGP neighbours have

performed their initial exchange of their complete routing databases, they only exchange updates to that information.

Routing Table Size

The Border Gateway Protocol (BGP) is used by BGP routers to route datagrams in the Internet Protocol (IP) network. According to this protocol, the various BGP routers exchange routing tables. Each router, adds to the routes stored in its own routing table, the routes it learns from his BGP neighbours, and then propagates this table to its neighbours. The routes are transmitted from one router to another. It results from this that the size of the routing table of each router can rapidly become very large. At the end the routing table can contain all routes known by every router participating to the Border Gateway Protocol (BGP) on the Internet : this table is called "the full Internet table". With the rapid expansion of Internet, the size of the “full internet table” has grown very rapidly. The size of the memory size required to store such a table in a router and the data processing capacity required to manage it, can represent a real problem. Some articles have been published pointing out the problem raised by this expansion. For instance, article entitled "Faster 'Net growth rate raises fears about routers" by Carolyn Duffy Marsan - Network World 04/02/01 - (http://www.nwfusion.com/news/2001/0402routing.html) indicates that the Internet is growing - in size and complexity - at a faster rate than today's routers can handle. After years of predictable growth, the size of the routing table and traffic is exploding.

To cope with the size increase of routing tables, manufacturers have continuously increased the power available in their routers in term of memory for storing said tables, and in term of data processing for consulting and updating said tables. In fact, network designers are faced with the following choices :

- either developing more powerful (and more expensive) routers in function of the size of the routing tables, or
- finding a way to artificially limit or decrease the size of the routing tables.

The usual technique for this last choice is to implement some filtering means and to throw away some of the table entries, for instance, entries related to small networks.

However this loss of information has some drawbacks, in particular, if the resulting routing is no more optimal.

With the increasing size of the internet table, some routers within networks can reach their limits. Suddenly, networks that were correctly operating, meet some problems related to the capacity of routers to route the traffic. In this case, a first solution is to remove old routers and to replace them by more powerful ones. Another solution is to artificially reduce the size of the full internet table by throwing away some routes and by implementing some filtering. In this last case, the choice of some routes may not be optimal anymore.

It is an object of the present invention, when the size of the routing table exceeds the storing and processing capacity of a BGP router, to replace said BGP router by a group of routers, without it is necessary that each of said routers has the capacity to manage alone the entire routing table.

It is a further object of the present invention to split the entire routing table of a BGP router between a group of several routers acting in a cooperative way, each router within the group, storing a portion of the entire routing table. It results that the size of the routing table stored in each router is significantly reduced, solving the problem of memory size and processing power required in each router for routing the traffic. When a router of the group receives a datagram and has to find a route to a destination which is unknown in its routing table, it forwards the datagram to the router within the group which is responsible for the portion of the routing table comprising the route for the destination. Thus an optimal routing is maintained at the price of an extra hop towards the router within the group in charge of the appropriate portion of the routing table.

Summary of the invention

The present invention is defined by the method set out in claim 1, the system set out in claim 10 and the program set out in claim 11.

The present invention is directed to a method, system and computer program as defined in independent claims for splitting and sharing routing information between several routers within a group of several routers acting as a single border router in an Internet protocol (IP) network, each router comprising a routing table. The method, for use in a router of the group, comprises the steps of :

- selecting routes in the routing table of the router;
- requesting other routers of the group to replace in their routing table each selected route with the router as next hop;
- associating part or all of non selected routes, each one with another router of the group;
- removing and replacing in the routing table, each non selected route associated with a router of the group by the associated router as next hop.

Further embodiments of the invention are provided in the appended dependent claims.

The foregoing, together with other objects, features, and advantages of this invention can be better appreciated with reference to the following specification, claims and drawings.

Brief description of the drawings

The novel and inventive features believed characteristics of the invention are set forth in the appended claims. The invention itself, however, as well as a preferred mode of use, further objects and advantages thereof, will best be understood by reference to the following detailed description of an illustrative detailed embodiment when read in conjunction with the accompanying drawings, wherein :

- Figure 1 is a general view of several Autonomous Systems interconnected by means of routers using an Exterior Gateway Protocol (EGP) to exchange routing information according to prior art.

- Figure 2 is a view of a typical network where a Border Gateway Protocol (BGP) router acts as a gateway between a private IP network and the Internet (the "external world").

- Figure 3 is a view of a network where a group of routers act as a gateway between a private IP network and the Internet, routing information being split and shared between said group of routers according to the present invention.

- Figure 4 is a flow chart showing different steps of the method of splitting and sharing routing information between a group of routers acting a a single border router. More particularly, Figure 4 shows the steps performed by a BGP router initiating the split of the routing table according to the present invention.

- Figure 5 is a flow chart showing different steps of the method of splitting and sharing routing information between several routers acting a a single border router. More particularly, Figure 5 shows the steps performed by a BGP router participating to the split of the routing table according to the present invention.

- Figure 6 shows the exchanges of messages between two routers sharing a Border Gateway Protocol (BGP) routing tables according to the present invention.

Preferred embodiment of the invention

PRIOR ART

Before going into the details of specific embodiments, it will be helpful to understand from a more general perspective the various elements and methods which may be related to the present invention.

Figure 1 is a general view of a network comprising several Autonomous Systems AS (100, 101, 102, 103) interconnected via border routers (104, 105, 106, 107). Said border routers exchange routing information related to the different Autonomous

Systems using the Exterior Gateway Protocol (EGP) (108). Today, the protocol the most commonly used between the public IP network (Internet) and private IP networks considered as Autonomous Systems (AS), is BGP (Border Gateway Protocol).

Figure 2 shows a typical situation for illustrating the problem generated by large routing tables in Border Gateway Protocol (BGP) routers. A private IP network (AS 0) (200), administered by a private company or a service provider, is connected (201) to other IP networks (AS 1, AS 2, AS 3) by means of a BGP router (200) acting as gateway. As shown in Figure 2, BGP router R0 (202) connects several BGP routers (203) (RI1, RI2 etc...). These BGP routers (203) are considered by router R0 (202) as BGP neighbours. BGP sessions are established between BGP router R0 and the BGP neighbours, and the routing tables are exchanged. When router R0 (202) experiences problems with the size of the routing table, the most common solution is to replace this router R0 by a more powerful BGP router.

INVENTION

Group of Routers Acting as a Single BGP Router

Instead of replacing BGP router R0 (202) by a more powerful router, the present invention discloses a method and system for using several routers together as a group. As shown in Figure 3, BGP router R0 (302) is replaced by a plurality of BGP routers (304) (R1,R2,R3,R4) forming a group. These BGP routers cooperate together by sharing their routing tables in a way to perform the same function as BGP router R0 (202) in Figure 2. The joint action of the BGP routers (304) of the group, allow them to access to the same routing information as BGP router R0, even if none of them, considered separately, is able to handle the entire routing table.

Initialisation of the Group

- First, each router member of the group (304) of BGP routers participating to the present invention, has to know the IP address of all other members of the group.
- In a first step, each BGP router of the group (304) establishes a BGP session (305) (also called EBGP session) through Internet with each BGP neighbours

(303) (RI1, RI2, RI3). In example described in Figure 3, BGP router R1 (304) establishes a EBGP session (305) with each of its BGP neighbours (303) (RI1, RI2, RI3).

- At initialisation time, each router of the group (304) establishes an IBGP session (306) with the other members of the group. Each BGP router of the group keeps an ordered list of active members of the group (304) (including itself). The method of ordering this list must be the same in all BGP routers within the group. However different methods using different criteria can be used to sort the list. A simple option is to order the list by ascending IP address.

If during the normal operations, one of the established BGP session is lost, or, if a BGP session, which initially failed, is finally established:

- all BGP sessions - including the EBGP (306) sessions with the BGP neighbours (303) (RI1, RI2, RI3) - are reset. The routing tables are cleared from any information related to the sessions. All sessions are reinitialized and reestablished according to the process previously described. The list of active BGP routers in each BGP router of the group is reordered.
- At the beginning, no route is advertised on the IBGP sessions between the different members of the group (304). Each member monitors itself the size of its own routing table :
  - As long as the size of the routing table remains below a predefined threshold, no particular action is taken.
  - When the size of the routing table goes beyond a predefined threshold, the BGP router decides to reduce the size of its routing table according to the process described here after.

Splitting and Sharing of Routing Information

The reduction of the size of the routing table of a BGP router, is based on the following principle : the entire routing table is split according to several subnets or subnetworks and shared between the different BGP routers of the group. A BGP router within the group, may decide to take the responsibility of routing the IP traffic intended to one of these subnetworks. The BGP router informs the other members of the group that it is ready to receive from them the IP traffic directed to this subnetwork. In consequence, all other BGP routers of the group can remove from their own routing table, the routes related to this subnetwork. These routes are

replaced by a single route pointing to the BGP router within the group, in charge of this subnetwork. The process can be repeated and each router can become responsible for one or a plurality of subnetworks.

Below is an example illustrating the method and system according to the present invention. To better understand and to make easier the explanation, a mask of /20 will be used. This parameter is configurable. In this particular example, the full Internet table is shared between the four BGP routers (R1, R2, R3, R4) of the group acting as a single BGP router (naturally, it is possible to use another number of routers). The IP addresses of these routers are in this same order. The claimed method comprises the following steps:

- **A. Initialisation of the Group of Routers :**
  - (400) At initialisation time, each BGP routers (R1, R2, R3 R4) of the group (304) establishes a EBGP session with its BGP neighbours (303) (BGP routers RI1, RI2, RI3).
  - (401) Each BGP router establishes a IBGP session with all other BGP routers of the group. Each BGP router of the group has been configured with the IP address of the other BGP routers of the group that it will share the routing tables with.
  - (402) At the end of this session establishment, each BGP router of the group builds a list with the active BGP routers of the group and orders this list (for instance by ascending IP address). In the present example, it is assumed that all BGP routers of the group are active and have successfully established a IBGP session with other BGP routers of the group. At the end of this step, each BGP router has build a list comprising routers “R1, R2, R3, R4”. At that time, no route is advertised on the IBGP sessions established between the different BGP routers of the group.

- **B. Split of the Routing Table of a Router between the other Routers of the Group :**
  - (403) When a BGP router of the group, for instance BGP router R1, detects that its routing table begins to exceed its storage and processing capabilities

(when, for instance, the size of its routing table exceeds a predefined threshold), this BGP router :

- (404) scans its routing table searching for a /20 subnet comprising a large number of routes pointing to some networks smaller than /20. BGP Router R1 splits the /20 subnet into four /22 subnet and decides to take the responsibility of one of the four /22 subnets (the subnet corresponding to its order in the list previously built). In the present example, because router R1 is the first router in the list, it will take the responsability of subnet number 1. If the /20 subnet selected by R1, starts with the address 1.1.0.0, then the four /22 subnets after the split of the routing table will start with addresses 1.1.0.0, 1.1.4.0, 1.1.8.0 and 1.1.12.0. Router R1 will take the responsibility of the /22 subnet starting with address 1.1.0.0. (the first one).
- (405) informs other BGP routers of the group (R2, R3, R4) by means of the previously established IBGP sessions that it takes the responsibility of a /22 subnet. The other BGP routers of the group update their routing table by replacing routes related to said subnet by a single route pointing to said subnet but with the BGP router R1 as next hop. In the present example the route will be :
  - network address 1.1.0.0,
  - network mask /22,
  - next hop router R1.

- **C. Sharing of the Routing Table between the Other Routers :**
  - (400) to (402) Each BGP router of the group (routers R2, R3, R4) is initialised according to the process previously described.
  - (500) Each BGP router of the group (routers R2, R3, R4) is invited by BGP router R1 via a message on previously established IBGP sessions, to update its routing table with a new route comprising BGP router R1 as next hop. At the receipt of this message, each BGP router (routers R2, R3, R4, R3) prepares itself to send IP traffic to BGP router R1.
  - (501) BGP router R2 removes from its routing table all routes under the responsibility of BGP router R1 and replace them by the new route pointing to BGP router R1. This operation allows to reduce the size of the routing table

stored in BGP router R2. In the present example, BGP router R2 removes all routes included in the /22 subnet starting with 1.1.0.0, and adds the route:

- network address 1.1.0.0,
- network mask /22, and
- next hop router R1.

The same process applies to BGP routers R3 and R4.

- (503) BGP router R2 computes the /20 subnet including the /22 subnet for which the new route has been received, and computes the four /22 subnets that are part of it. These four /22 subnets are :
  - 1 : mask /22, address 1.1.0.0.
  - 2 : mask /22, address 1.1.4.0.
  - 3 : mask /22, address 1.1.8.0.
  - 4 : mask /22, address 1.1.12.0.

  BGP router R2 assigns itself the responsibility of one of the /22 subnet included into the /20 (the subnet corresponding to the order of its address in the list previously built). Because BGP router R2 is the second in the router list, it will take the responsibility of subnet 2. In the present example, the /22 subnet starting with 1.1.0.0 is included into a larger /20 subnet also starting with 1.1.0.0. The BGP router R2 then splits this larger subnet into four /22 subnets and assigns itself the second subnet starting with 1.1.4.0.
- (504) BGP router R2 now behaves exactly as router R1 : it informs all other routers in the group using the IBGP sessions, that now a single route with it as next hop, is pointing to the /22 subnet it has the responsibility. In the present example, BGP router R2 informs the other routers of the group that it of the following route :
  - network address 1.1.4.0,
  - network mask /22, and
  - next hop router R2.

Process Convergence

Each BGP routers of the group behaves the same way : each time a route is received from one of the BGP routers of the group, it :

- replace all routes comprised in the /22 subnet into a single route pointing to this router,

- take the responsability of a /22 subnet, and
- sends the corresponding route to the other BGP routers of the group.

At the end, each BGP router of the group :

- is responsible of one of the /22 subnets, and
- has received from each of the other BGP routers of the group, one route for each of the other /22 subnets comprised in the initial /20 subnet.
- has removed all the routes corresponding to each /22 subnet it is not responsible for.
- has replaced the routes previously removed by a single route.

If one of the BGP routers experiences problems with the size of its routing table, this router will initiate again the routing table reduction process described here above with another subnet.

Figure 6 shows how two BGP routers within the group interact in order to split and share their routing table.

- (600) As described earlier, BGP router R1 initiates the process when the size of its routing table exceeds its storage and processing capabilities (for instance, when the size of its routing table exceeds a given threshold). Router R1 assigns itself a subnetwork and informs the other BGP routers of the group to update their routing table in order to route towards it the IP traffic intended to this subnetwork. In our particular example, router R1 sends to router R2 a message comprising information related to a new route :
  - network address 1.1.0.0;
  - network mask /22;
  - next hop router R1.
- (601) At receipt of this message, router R2 updates its routing table. In the present example, all routes including subnetwork 1.1.0.0 /22 are removed and replaced by a single route with router R1 as next hop. Then router R1 assigns itself a /22 subnetwork. Finally, router R2 sends to the other routers of the group, including router R1, a message with a new route:
  - network address 1.1.4.0;
  - network mask /22:
  - next hop R2.

- (602) At reception of the message sent by R2, router R1 updates its routing table the same way as described previously for router R2. In the present example, all routes including subnetwork 1.1.4.0 /22 are removed and replaced by a single route with router R2 as next hop. Note that router R1 will receive a similar message from each BGP router of the group.

Recovery

The BGP router within the group detects the loss of a IBGP session (for instance, because a BGP router in the group has failed), the process is reinitialised. The BGP router resets all its IBGP sessions and rebuilds an ordered list with all the BGP routers participating to the group and which are sharing their routing tables. The process is the same when a BGP router of the group initially in failure, is recovered (when the IBGP session with this router is reestablished). It is important to note that the temporary absence or unavailability of a BGP router in the group does not prevent the process from working. The process is just less efficient.

EXAMPLE

The following is a more complete and detailed example to better illustrate the method according to the present invention.

At the beginning of the process, the BGP routers of the group (routers R1, R2, R3, R4) establish EBGP sessions with their BGP neighbours (routers RI1, RI2, RI3) and receive the same routing table. In the present example the received routing table comprises the following information :

Initial routing tables in BGP routers R1 R2 R3 R4 :

Network	Mask	Next Hop
1.1.0.0	255.255.240.0 (/20)	RI1
1.1.0.0	255.255.255.0 (/24)	RI2
1.1.1.0	255.255.255.0 (/24)	RI3
1.1.2.0	255.255.255.0 (/24)	RI2
1.1.4.0	255.255.252.0 (/24)	RI3
1.1.5.0	255.255.252.0 (/24)	RI2
1.1.6.0	255.255.252.0 (/24)	RI1
1.1.8.0	255.255.252.0 (/24)	RI3

1.1.9.0	255.255.252.0 (/24)	RI1
1.1.10.0	255.255.252.0 (/24)	RI2
1.1.12.0	255.255.252.0 (/24)	RI1
1.1.13.0	255.255.252.0 (/24)	RI3
1.1.14.0	255.255.252.0 (/24)	RI2
1.1.16.0	255.255.252.0 (/20)	RI2
1.1.32.0	255.255.252.0 (/20)	RI3

BGP Routers R1 R2 R3 R4 also establish IBGP session between each other. At the beginning of the process, no information is advertised on these IBGP sessions. In our example, BGP routers R1, R2, R3, R4, have their IP address in the same order.

When BGP router R1 decides to reduce its routing table for problems of storage or processing capacity (or for any other problem), it first selects a /20 subnet comprising several routes. In the present example, BGP router R1 selects subnet 1.1.0.0 /20 and, using the IBGP sessions, announce to the other BGP routers of the group (R2, R3, R4), the following route :

- Network : 1.1.0.0
- Mask /22
- Next hop R1.

At receipt of this route, BGP router R2 assigns to itself a second /22 subnet 1.1.4.0 /22 and start to reduce its routing table based on the route received from the BGP router R1. The routing table of BGP router R2 becomes as follows :

Routing table of BGP router R2 :

Network	Mask	Next Hop
1.1.0.0	255.255.240.0 (/20)	RI1
1.1.0.0	255.255.255.0 (/22)	**R1**
1.1.4.0	255.255.252.0 (/24)	RI3
1.1.5.0	255.255.252.0 (/24)	RI2
1.1.6.0	255.255.252.0 (/24)	RI1
1.1.8.0	255.255.252.0 (/24)	RI3
1.1.9.0	255.255.252.0 (/24)	RI1
1.1.10.0	255.255.252.0 (/24)	RI2
1.1.12.0	255.255.252.0 (/24)	RI1
1.1.13.0	255.255.252.0 (/24)	RI3
1.1.14.0	255.255.252.0 (/24)	RI2
1.1.16.0	255.255.252.0 (/20)	RI2
1.1.32.0	255.255.252.0 (/20)	RI3

BGP Router R2 sends to the other BGP routers of the group (routers R1, R3, R4) using the IBGP sessions, the following route :

- Network : 1.1.4.0,
- Mask /22,
- Next hop router R2.

At receipt of this route, BGP router R1 also proceeds with the reduction of the routing table. The routing table of BGP router R1 becomes as follows :

Routing table of BGP router R1 :

Network	Mask	Next Hop
1.1.0.0	255.255.240.0 (/20)	RI1
1.1.0.0	255.255.255.0 (/24)	RI2
1.1.1.0	255.255.255.0 (/24)	RI3
1.1.2.0	255.255.255.0 (/24)	RI2
1.1.4.0	255.255.255.0 (/22)	**R2**
1.1.8.0	255.255.252.0 (/24)	RI3
1.1.9.0	255.255.252.0 (/24)	RI1
1.1.10.0	255.255.252.0 (/24)	RI2
1.1.12.0	255.255.252.0 (/24)	RI1
1.1.13.0	255.255.252.0 (/24)	RI3
1.1.14.0	255.255.252.0 (/24)	RI2
1.1.16.0	255.255.252.0 (/20)	RI2
1.1.32.0	255.255.252.0 (/20)	RI3

In the meantime BGP routers R3 and R4, receive the routes sent by BGP routers R1 and R2 and starts the reduction of their routing table. They send to the other BGP routers of the group, the routes pointing to the subnets they take in charge, respectively route :

- Network : 1.1.8.0,
- Mask /22,
- Next hop R3,

and route :

- Network : 1.1.12.0,
- Mask /22,
- Next hop R4.

The process is converging and at the end, we find the following tables :

Final routing table of router R1 :

Network	Mask	Next Hop
1.1.0.0	255.255.240.0 (/20)	RI1
1.1.0.0	255.255.255.0 (/24)	RI2
1.1.1.0	255.255.255.0 (/24)	RI3
1.1.2.0	255.255.255.0 (/24)	RI2
1.1.4.0	255.255.255.0 (/22)	**R2**
1.1.8.0	255.255.255.0 (/22)	**R3**
1.1.12.0	255.255.255.0 (/22)	**R4**
1.1.16.0	255.255.252.0 (/20)	RI2
1.1.32.0	255.255.252.0 (/20)	RI3

Final routing table of router R2 :

Network	Mask	Next Hop
1.1.0.0	255.255.240.0 (/20)	RI1
1.1.0.0	255.255.255.0 (/22)	**R1**
1.1.4.0	255.255.252.0 (/24)	RI3
1.1.5.0	255.255.252.0 (/24)	RI2
1.1.6.0	255.255.252.0 (/24)	RI1
1.1.8.0	255.255.255.0 (/22)	**R3**
1.1.12.0	255.255.255.0 (/22)	**R4**
1.1.16.0	255.255.252.0 (/20)	RI2
1.1.32.0	255.255.252.0 (/20)	RI3

5 Final routing table of router R3 :

Network	Mask	Next Hop
1.1.0.0	255.255.240.0 (/20)	RI1
1.1.0.0	255.255.255.0 (/22)	**R1**
1.1.4.0	255.255.255.0 (/22)	**R2**
1.1.8.0	255.255.252.0 (/24)	RI3
1.1.9.0	255.255.252.0 (/24)	RI1
1.1.10.0	255.255.252.0 (/24)	RI2
1.1.12.0	255.255.255.0 (/22)	**R4**
1.1.16.0	255.255.252.0 (/20)	RI2
1.1.32.0	255.255.252.0 (/20)	RI3

Final routing table of router R4 :

Network	Mask	Next Hop
1.1.0.0	255.255.240.0 (/20)	RI1
1.1.0.0	255.255.255.0 (/22)	**R1**

1.1.4.0	255.255.255.0 (/22)	**R2**
1.1.8.0	255.255.255.0 (/22)	**R3**
1.1.12.0	255.255.252.0 (/24)	RI1
1.1.13.0	255.255.252.0 (/24)	RI3
1.1.14.0	255.255.252.0 (/24)	RI2
1.1.16.0	255.255.252.0 (/20)	RI2
1.1.32.0	255.255.252.0 (/20)	RI3

While the invention has been particularly shown and described with reference to a preferred embodiment, it will be understood that various changes in form and detail may be made therein without departing from the spirit, and scope of the invention.

Claims

What we claim is:

1. A method for splitting and sharing routing information between several routers within a group of several routers acting as a single border router in an Internet protocol (IP) network, each router comprising a routing table, said method, for use in a router of the group, comprising the steps of :

selecting routes in the routing table of the router;

- requesting other routers of the group to replace in their routing table each selected route with the router as next hop;
- associating part or all of non selected routes, each one with another router of the group;
- removing and replacing in the routing table, each non selected route associated with a router of the group by the associated router as next hop.

2. The method according to the preceding claim comprising the further steps of :

- forwarding IP traffic corresponding to a non selected route, to the router associated with said route within the routing table.

3. The method according to any one of the preceding claims wherein the step of requesting other routers of the group to replace in their routing table each selected route with the router as next hop, comprises the further step of :

- receiving from other routers the IP traffic corresponding to the selected routes;
- routing said IP traffic.

4. The method according to any one of the preceding claims wherein the step of associating part or all of non selected routes, each one with another router of the group, comprises the steps of :

- receiving from each other router of the group means for associating part or all of non selected routes, each one with another router of the group.

5. The method according to the preceding claim wherein the step of selecting routes in the routing table comprises the further step of :

- selecting contiguous IP addresses within a given address range.

6. The method according to any one of the preceding claims comprising the preliminary steps of :

- establishing sessions with other routers of the group;
- creating a list of routers of the group.

7. The method according to any one of the preceding claims comprising the preliminary steps of :

- establishing sessions with other border routers.

8. The method according to any one of the preceding claims wherein the step of selecting routes comprises the preliminary step of :

- comparing the size of the routing table with a predefined threshold.

9. The method according to any one of the preceding claims wherein routers within the group are exchanging routing information using the Border Gateway Protocol (BGP).

10. A router comprising means adapted for carrying out all the steps of the method according to any one of the preceding claims.

11. A computer program comprising instructions for carrying out the method according to any one of claims 1 to 9 when said computer program is executed on system according to claim 10.

METHOD AND SYSTEM FOR SPLITTING AND SHARING ROUTING INFORMATION BETWEEN SEVERAL ROUTERS ACTING AS A SINGLE BORDER ROUTER

Abstract

The present invention is directed to a method, system and computer program for splitting and sharing routing information between several routers within a group of several routers acting as a single border router in an Internet protocol (IP) network, each router comprising a routing table. The method, for use in a router of the group, comprises the steps of :

selecting routes in the routing table of the router;

- requesting other routers of the group to replace in their routing table each selected route with the router as next hop;
- associating part or all of non selected routes, each one with another router of the group;
- removing and replacing in the routing table, each non selected route associated with a router of the group by the associated router as next hop.

Figure 3

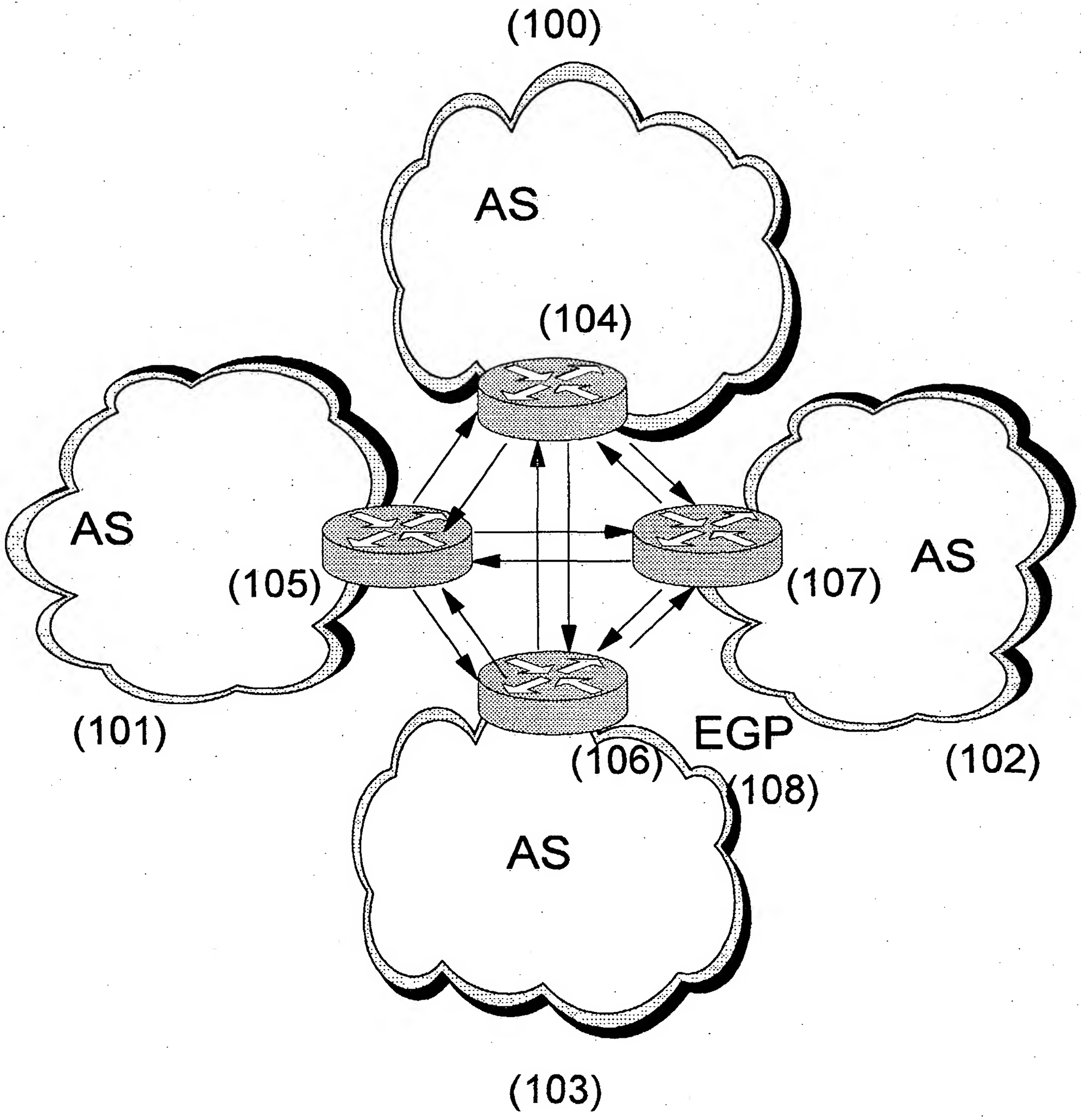


Figure 1 : interconnected Autonomous Systems

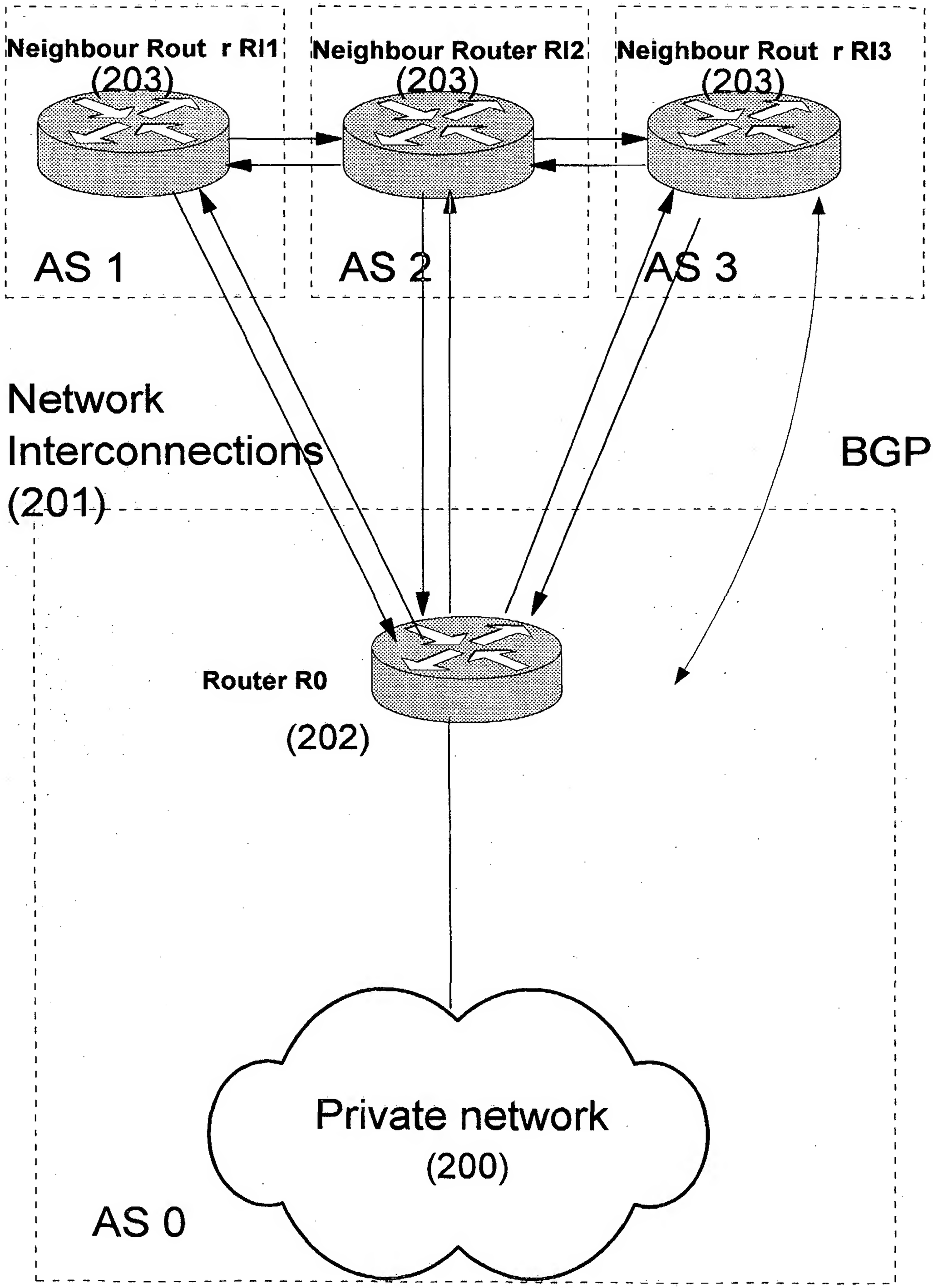
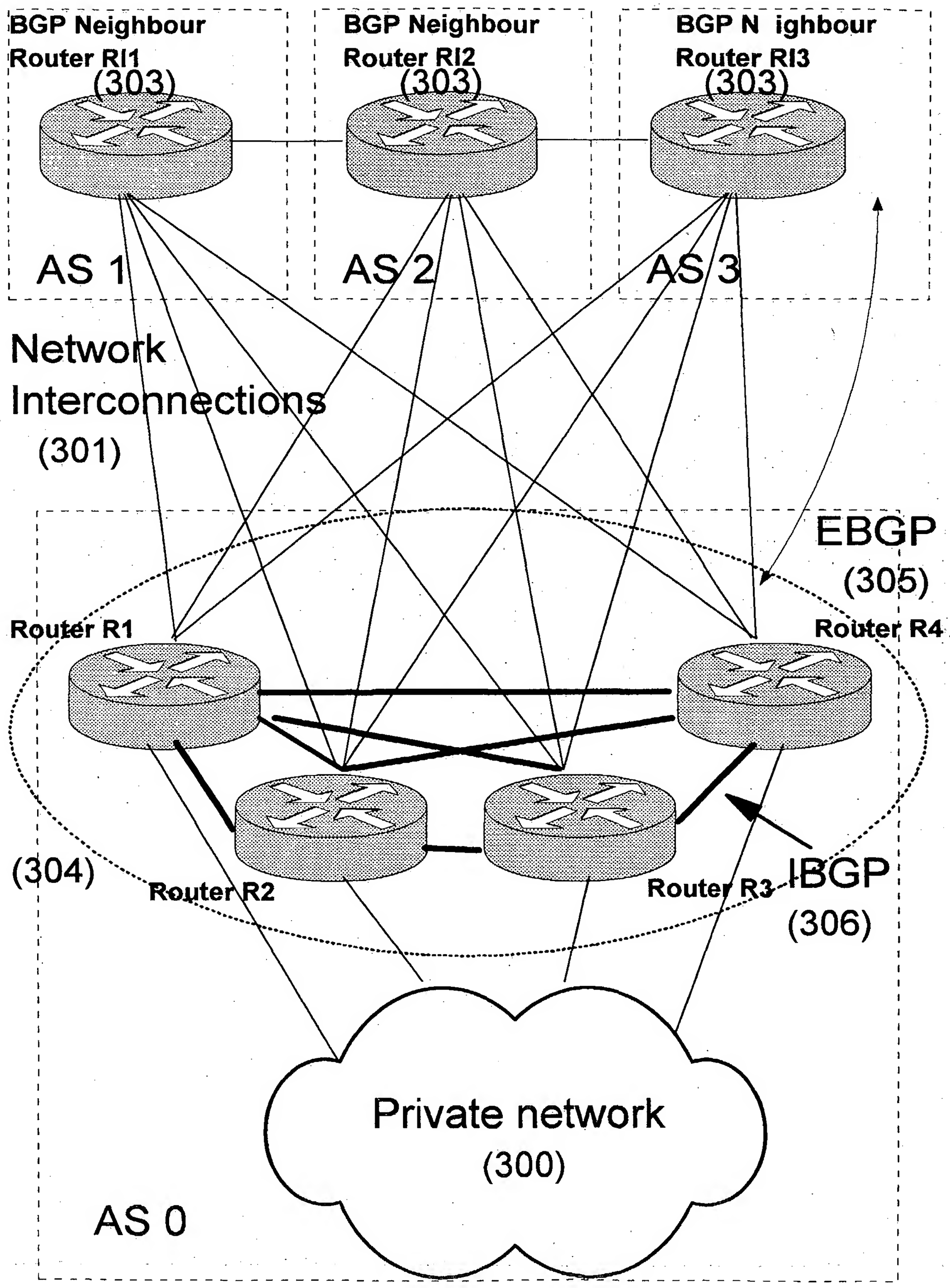


Figure 2 : BGP router acting as gateway


BGP Neighbour Router RI1
(303)
BGP Neighbour Router RI2
(303)
BGP N ighbour Router RI3
(303)
AS 1
AS 2
AS 3
Network Interconnections (301)
EBGP (305)
Router R1
Router R4
Router R2
Router R3
IBGP (306)
(304)
Private network (300)
AS 0

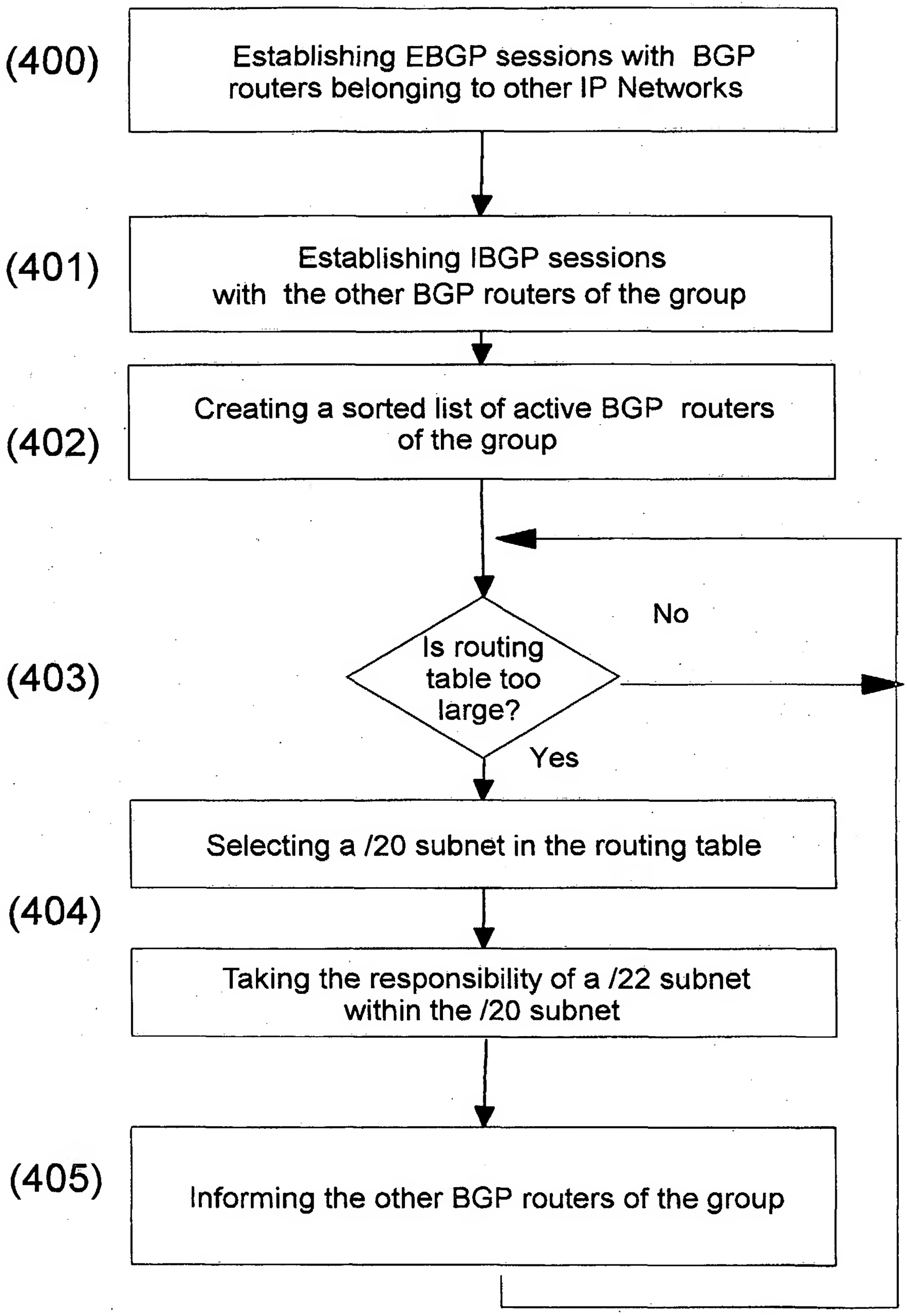
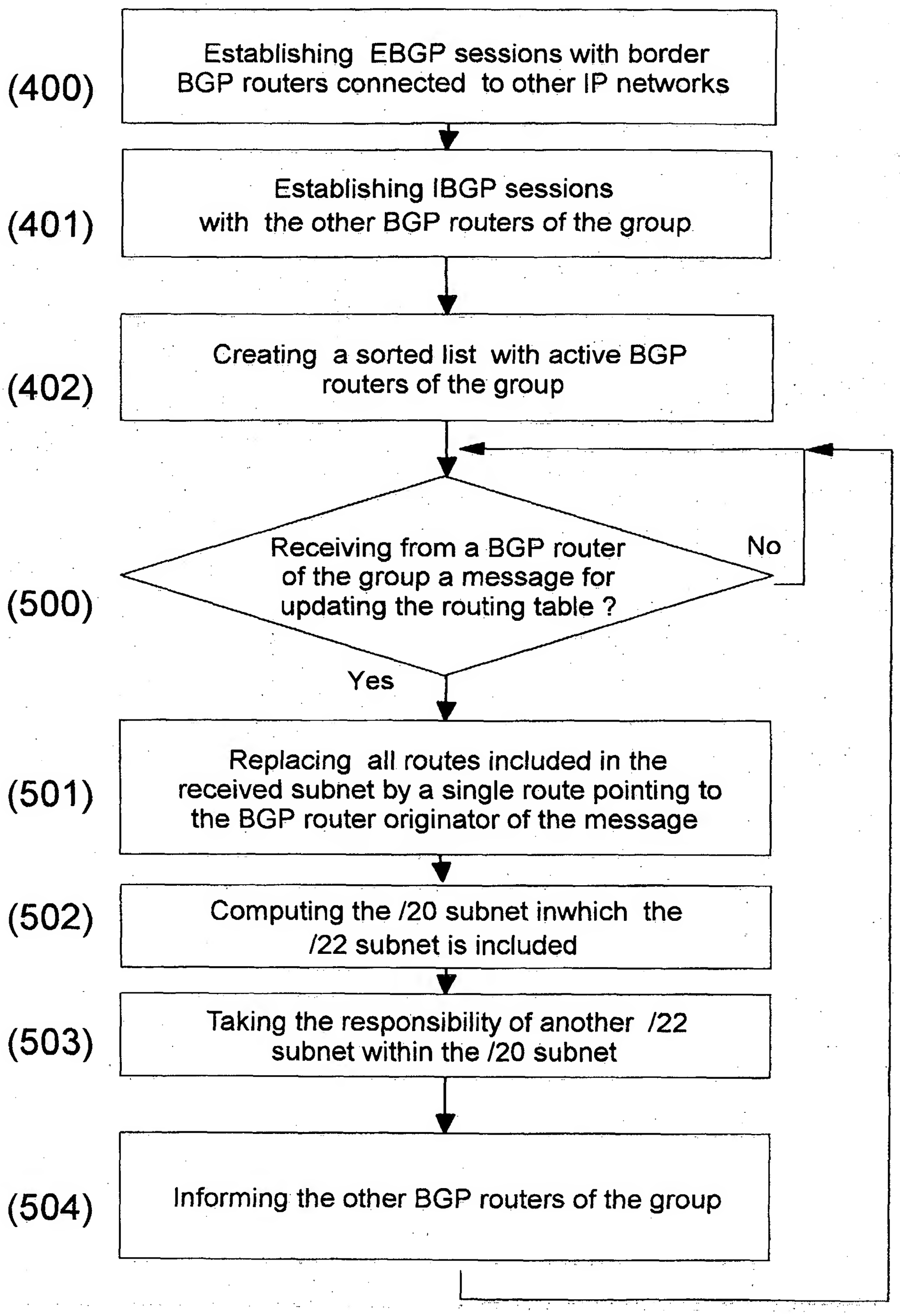


Figure 4 : BGP router initialization and routing table split


(400)
Establishing EBGP sessions with border BGP routers connected to other IP networks
(401)
Establishing IBGP sessions with the other BGP routers of the group
(402)
Creating a sorted list with active BGP routers of the group
(500)
Receiving from a BGP router of the group a message for updating the routing table ?
No
Yes
(501)
Replacing all routes included in the received subnet by a single route pointing to the BGP router originator of the message
(502)
Computing the /20 subnet inwhich the /22 subnet is included
(503)
Taking the responsibility of another /22 subnet within the /20 subnet
(504)
Informing the other BGP routers of the group

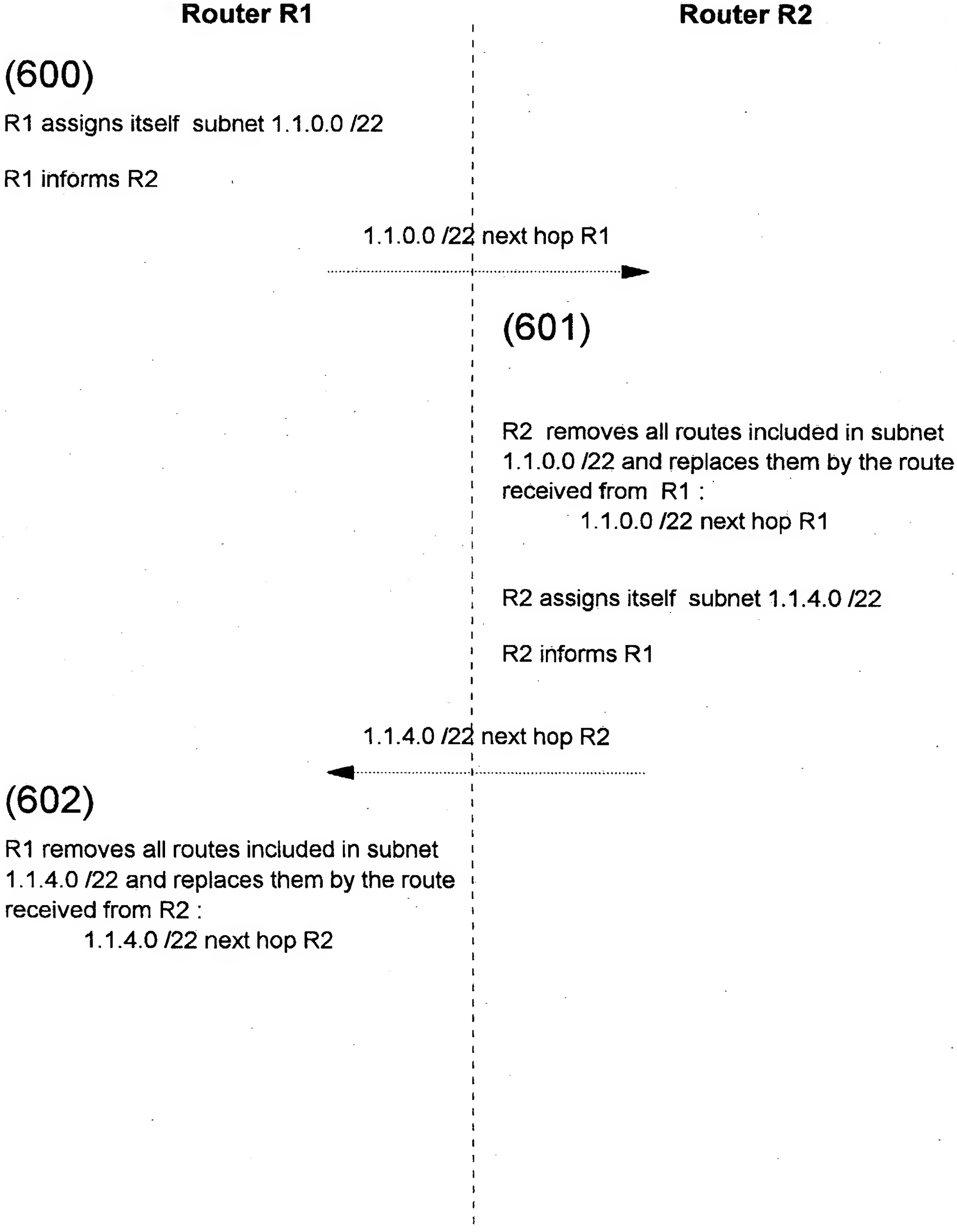


Figure 6 : message exchange between BGP routers of a group